MODULE-3
DESIGN OF SAMPLE SURVEYS

Strongly agree
Agree
Disagree
ly disagree

INTRODUCTION

- All items in any field of inquiry constitute a 'Universe' or 'Population.'
- A complete enumeration of all items in the 'population' is known as a census inquiry or census survey.
- It is obvious that for any study or investigation census survey is rather infeasible.
- For example, to have an idea of average per capita monthly income of the people in India, we will have to enumerate all the earning individuals in the country, which is a very difficult task.

- Census survey is impossible in the situations when population is infinite.
- In some cases when population is finite, but the units are destroyed while inspected, census survey is not at all desirable, e.g ., inspection of crackers.
- Further, many a time it is not possible to examine every item in the population, and sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population.
- In such cases there is no utility of census surveys.

- The selected respondents constitute what is technically called a 'sample' and the selection process is called 'sampling technique.' The survey so conducted is known as 'sample survey'.
- Algebraically, let the population size be N and if a part of size n (which is < N) of this population is selected according to some rule for studying some characteristic of the population, the group consisting of these n units is known as 'sample'.
- Researcher must prepare a sample design for his study i.e ., he must plan how a sample should be selected and of what size such a sample would be.

SAMPLE DESIGN

- A sample design is a definite plan for obtaining a sample from a given population.
- It refers to the technique or the procedure the researcher would adopt in selecting items for the sample.
- Sample design may as well lay down the number of items to be included in the sample i.e ., the size of the sample.

SAMPLE DESIGN

- Sample design is determined before data are collected.
- There are many sample designs which a researcher can choose.
- Some designs are relatively more precise and easier to apply than others.
- Researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

The main steps of sampling design are as follows:

1) **Objective:** The first step of sampling design is to define the objectives of survey in clear and concrete terms. The sponsors or the researchers of the survey should confirm that the objectives are commensurate with the money, manpower and time limit available for the survey.

2) **Population:** In order to meet the objectives of the survey, what should be the population? This question should be answered in the second step. The population should be clearly defined.

3) **Sampling units and frame:** A decision has to be taken concerning a sampling unit before selecting sample.

- Sampling unit may be a geographical one such as state, district, village, etc ., or a construction unit such as house, flat, etc., or it may be a social unit such as family, club, school, etc ., or it may be an individual.
- The researcher will have to decide one or more of such units that he has to select for his study. The list of sampling units is called as 'frame' or sampling frame. Sampling frame contains the names of all items of a universe (in case of finite universe only).
- Such a list should be comprehensive, correct, reliable and appropriate. It is extremely important for the source list to be as representative of the population as possible.

4) **Size of sample:** This refers to the number of items to be selected from the universe to constitute a sample. This is a major problem before a researcher.

- The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility.
- While deciding the size of sample, researcher must determine the desired precision as also an acceptable confidence level for the estimate. The size of population variance needs to be considered as in case of larger variance usually a bigger sample is needed.

- The size of population must be kept in view for this also limits the sample size. The parameters of interest in a research study must be kept in view, while deciding the size of the sample.
- Costs too dictate the size of sample that we can draw. As such, budgetary constraint must invariably be taken into consideration when we decide the sample size.

5)Parameters of interest:

Statistical constants of the population are called as parameters, e.g ., population mean, population proportion etc.

- When we do census survey we get the actual value of parameters.
- On the other hand, when we do sample survey we get the estimates of unknown population parameters in place of their actual values.

- In determining the sample design, one must consider the question of the specific population parameters which are of interest.
- For instance, we may be interested in estimating the proportion of persons with some characteristic in the population, or we may be interested in knowing some average or the other measure concerning the population.
- There may also be important sub-groups in the population about whom we would like to make estimates.
- All this has a strong impact upon the sample design we would accept.

6) Data collection: No irrelevant information should be collected, and no essential information should be discarded, the objectives of the study should be very much clear in the mind of surveyor.

7) Non-respondents: Because of practical difficulties, data may not be collected for all the sampled units. This non-response tends to change the results. The reason for non-response should be recorded by the investigator. Such cases should be handled with caution.

8) Selection of proper sampling design: The researcher must decide the type of sample he will use i.e., he must decide about the technique to be used in selecting the items for the sample. There are several sample designs out of which the researcher must chose one for his study.

9) Organizing field work: The success of a survey depends on the reliable field work. There should be efficient supervisory staff and trained professional for the field work.

10) Pilot survey: It is always helpful to try out the research design on a small scale before going to the field. This is called as pilot survey. It might give the better idea of practical problems and troubles.

11) Budgetary constraint: Cost considerations, from practical point of view, have a major impact upon decisions relating to not only the size of sample but also to the type of sample.

SAMPLING AND NON-SAMPLING ERRORS

- The errors involved in the collection of data are classified into sampling and non-sampling errors.

I) **Sampling error:** can be measured for a given sample design and size. The measurement of sampling error is usually called the 'precision of the sampling plan'.

- If we increase the sample size, the precision can be improved. But increasing the size of the sample has its own limitations viz., a large sized sample increases the cost of collecting data and also enhances the systematic bias.

- Thus, the effective way to increase precision is usually to select a better sampling design which has a smaller sampling error for a given sample size at a given cost.
- In practice, however, people prefer a less precise design because it is easier to adopt the same and also because of the fact that systematic bias can be controlled in a better way in such a design.
- In brief, while selecting a sampling procedure, researcher must ensure that the procedure causes a relatively small sampling error and helps to control the systematic bias in a better way.

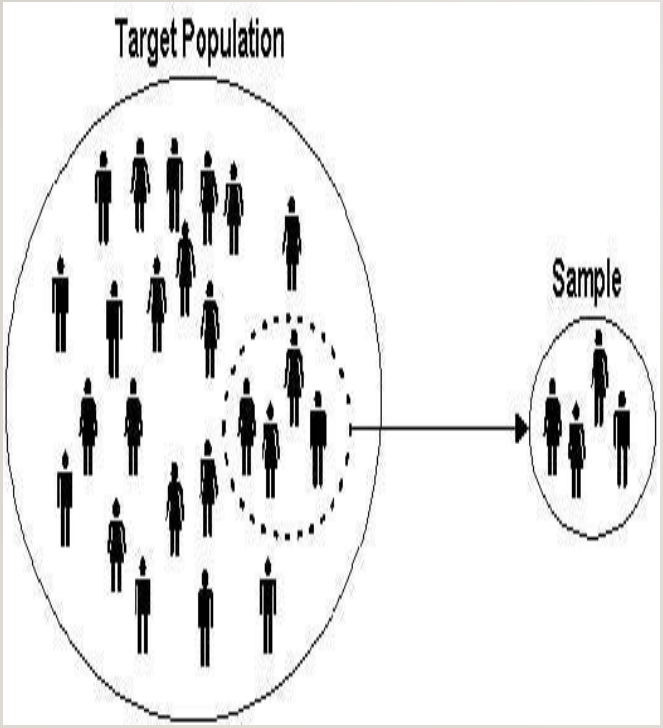
2) **Non-sampling errors:** Non-sampling errors arise at the stage of collection and preparation of data and thus are present in both the sample survey as well as the census survey.

- Thus the data obtained in census survey is free form sampling errors, however subjected to non-sampling errors.
- Non-sampling errors can be reduced by defining the sample units, frame and the population correctly and by employing efficient people in the investigations.

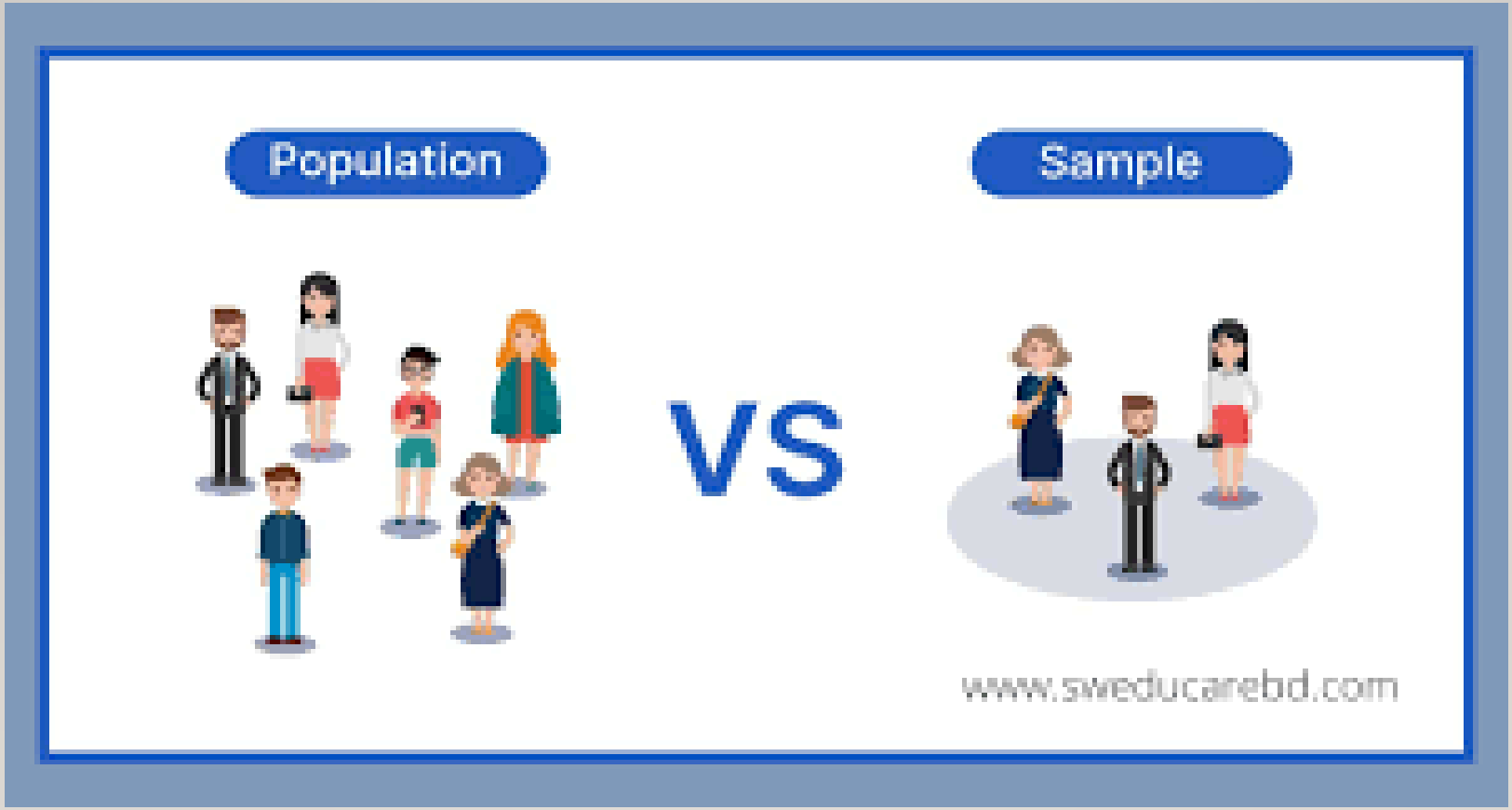
SAMPLE SURVEY V/S CENSUS SURVEY

- In a sample survey, since we study only a subpart of the whole population, requires less money and less time.
- Most of the times, non-sampling errors are so much large that the results of sample survey are much more accurate than those of census survey.
- Non-sampling errors arise due to a number of factors such as inefficiency of field workers, non-response, bias due to interviewer, etc.
- These errors are likely to grow when the number of units inspected increase.

SAMPLE SURVEY



CENSUS SURVEY


Population
VS
Sample
www.sweducarebd.com

DIFFERENT TYPES OF SAMPLE DESIGNS

The method of selecting a sample is of fundamental importance and depends on the nature of data and investigation. The techniques of selecting a sample are classified as- Non-probability sampling and probability sampling.

- **Non-probability sampling**: Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample.
- Non-probability sampling is also known by different names such as deliberate sampling, purposive sampling and judgement sampling.

- In this type of sampling, items for the sample are selected deliberately by the researcher, his choice concerning the items remains supreme
- In other words, under non-probability sampling the organizers of the inquiry purposively choose the particular units of the universe for constituting a sample on the basis that the small mass that they so select out of a huge one will be typical or representative of the whole.

- For instance, if economic conditions of people living in a state are to be studied, a few towns and villages may be purposively selected for intensive study on the principle that they can be representative of the entire state.
- Thus, the judgement of the organizers of the study plays an important part in this sampling design

Quota sampling - is also an example of non-probability sampling. Under quota sampling the interviewers are simply given quotas to be filled from the different strata, with some restrictions on how they are to be filled.

- In other words, the actual selection of the items for the sample is left to the interviewer's discretion.
- This type of sampling is very convenient and is relatively inexpensive.
- But the samples so selected certainly do not possess the characteristic of random samples.
- Quota samples are essentially judgement samples and inferences drawn on their basis are not amenable to statistical treatment in a formal way.

2) Probability sampling:

- Probability sampling is also known as 'random sampling' or 'chance sampling'.

- Under this sampling design, every item of the universe has an equal chance of inclusion in the sample.
- It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process.
- Here it is blind chance alone that determines whether one item or the other is selected.

- The results obtained from probability or random sampling can be assured in terms of probability i.e., we can measure the errors of estimation or the significance of results obtained from a random sample, and this fact brings out the superiority of random sampling design over the deliberate sampling design.
- Random sampling ensures the law of Statistical Regularity which states that if on an average the sample chosen is a random one, the sample will have the same composition and characteristics as the universe.
- This is the reason why random sampling is considered as the best technique of selecting a representative sample.

SIMPLE RANDOM SAMPLING

- Simple Random sampling from a finite population refers to that method of sample selection which gives each possible sample combination an equal probability of being picked up and each item in the entire population to have an equal chance of being included in the sample.
- This applies to sampling without replacement i.e., once an item is selected for the sample, it cannot appear in the sample again In brief, the implications of random sampling (or simple random sampling) are:

  (a) It gives each element in the population an equal probability of getting into the sample; and all choices are independent of one another.

  (b) It gives each possible sample combination an equal probability of being chosen.

- Keeping this in view we can define a simple random sample (or simply a random sample) from a finite population as a sample which is chosen in such a way that each of the NCn possible samples has the same probability, 1/NCn , of being selected.
- To make it more clear we take a certain finite population consisting of six elements (say a, b, c, d, e, f) i.e., N = 6. Suppose that we want to take a sample of size n = 3 from it. Then there are 6 C3 = 20 possible distinct samples of the required size, and they consist of the elements abc, abd, abe, abf, acd, ace, acf, ade, adf, aef, bcd, bce, bcf, bde, bdf, bef, cde, cdf, cef, and def.

- If we choose one of these samples in such a way that each has the probability 1/20 of being chosen, we will then call this a random sample
- We can illustrate the procedure by an example. First of all we reproduce the first thirty sets of Tippett's numbers

2952	6641	3992	9792	7979	5911
3170	5624	4167	9525	1545	1396
7203	5356	1300	2693	2370	7483
3408	2769	3563	6107	6913	7691
0560	5246	1112	9025	6008	8126

- Suppose we are interested in taking a sample of 10 units from a population of 5000 units, bearing numbers from 3001 to 8000. We shall select 10 such figures from the above random numbers which are not less than 3001 and not greater than 8000. If we randomly decide to read the table numbers from left to right, starting from the first row itself, we obtain the following numbers: 6641, 3992, 7979, 5911, 3170, 5624, 4167, 7203, 5356, and 7483.
- The units bearing the above serial numbers would then constitute our required random sample.

- One may note that it is easy to draw random samples from finite populations with the aid of random number tables only when lists are available, and items are readily numbered.
- But in some situations, it is often impossible to proceed in the way we have narrated above. For example, if we want to estimate the mean height of trees in a forest, it would not be possible to number the trees, and choose random numbers to select a random sample.
- In such situations what we should do is to select some trees for the sample haphazardly without aim or purpose, and should treat the sample as a random sample for study purposes.

COMPLEX RANDOM SAMPLING DESIGNS:

- Some complex random sampling designs, which are mixtures of probability and non-probability sampling methods as below.

(i) Systematic sampling: In some instances, the most practical way of sampling is to select every ith item on a list. Sampling of this type is known as systematic sampling.

- An element of randomness is introduced into this kind of sampling by using random numbers to pick up the unit with which to start.

Systematic Sampling

[ˌsi-stə-ˈma-tik ˈsam-pliŋ]

A method for selecting a sample from a population in a randomized manner, so as to minimize sampling error.

Investopedia

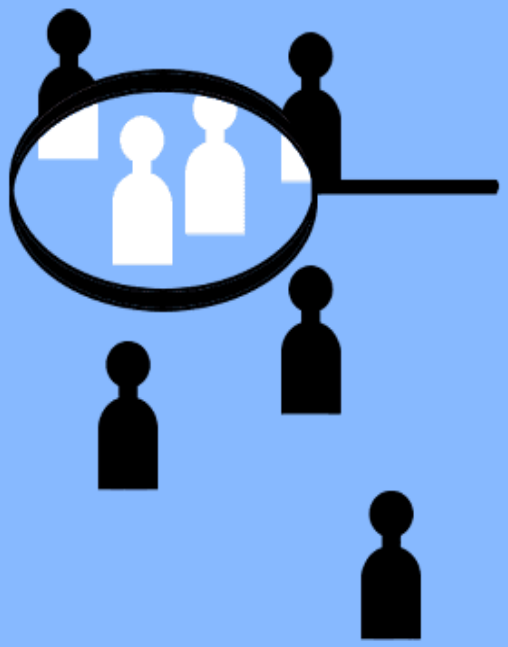
- For instance, if a 4 per cent sample is desired, the first item would be selected randomly from the first twenty-five and thereafter every 25th item would automatically be included in the sample.
- Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals.

- Although a systematic sample is not a random sample in the strict sense of the term, but it is often considered reasonable to treat systematic sample as if it were a random sample
- Systematic sampling has certain plus points.
- It can be taken as an improvement over a simple random sample in as much as the systematic sample is spread more evenly over the entire population.
- It is an easier and less costlier method of sampling and can be conveniently used even in case of large populations.
- But there are certain dangers too in using this type of sampling. If there is a hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling.

- For instance, every 25th item produced by a certain production process is defective. If we are to select a 4% sample of the items of this process in a systematic manner, we would either get all defective items or all good items in our sample depending upon the random starting position.
- If all elements of the universe are ordered in a manner representative of the total population, i.e., the population list is in random order, systematic sampling is considered equivalent to random sampling.

- But if this is not so, then the results of such sampling may, at times, not be very reliable. In practice, systematic sampling is used when lists of population are available and they are of considerable length.

(ii)Stratified sampling: If a population from which a sample is to be drawn does not constitute a homogeneous group, stratified sampling technique is generally applied in order to obtain a representative sample.



Stratified Random Sampling

['stra-tə-,fīd 'ran-dəm 'sam-pliŋ]

A method of sampling that involves the division of a population into smaller sub-groups known as strata.

Investopedia

- Under stratified sampling the population is divided into several sub-populations that are individually more homogeneous than the total population (the different sub-populations are called 'strata') and then we select items from each stratum to constitute a sample.
- Since each stratum is more homogeneous than the total population, we are able to get more precise estimates for each stratum and by estimating more accurately each of the component parts, we get a better estimate of the whole.
- In brief, stratified sampling results in more reliable and detailed information.

The following three questions are highly relevant in the context of stratified sampling:

(a) How to form strata?

(b) How should items be selected from each stratum?

(c) How many items be selected from each stratum or how to allocate the sample size of each stratum?

- Regarding the first question, we can say that the strata be formed on the basis of common characteristic(s) of the items to be put in each stratum.
- This means that various strata be formed in such a way as to ensure elements being most homogeneous within each stratum and most heterogeneous between the different strata.

- In respect of the second question, we can say that the usual method, for selection of items for the sample from each stratum, resorted to is that of simple random sampling. Systematic sampling can be used if it is considered more appropriate in certain situations.
- Regarding the third question, we usually follow the method of proportional allocation under which the sizes of the samples from the different strata are kept proportional to the sizes of the strata.

- In cases where strata differ not only in size but also in variability and it is considered reasonable to take larger samples from the more variable strata and smaller samples from the less variable strata, we can then account for both (differences in stratum size and differences in stratum variability) by using disproportionate sampling design by requiring:

$$n_1/N_1\sigma_1 = n_2/N_2\sigma_2 = \ldots\ldots\ldots = n_k/N_k\sigma_k$$

where $\sigma_1, \sigma_2, \ldots$ and σ_k denote the standard deviations of the k strata, $N_1, N_2, \ldots, N_k$ denote the sizes of the k strata and $n_1, n_2, \ldots, n_k$ denote the sample sizes of k strata. This is called '*optimum allocation*' in the context of disproportionate sampling. The allocation in such a situation results in the following formula for determining the sample sizes different strata:

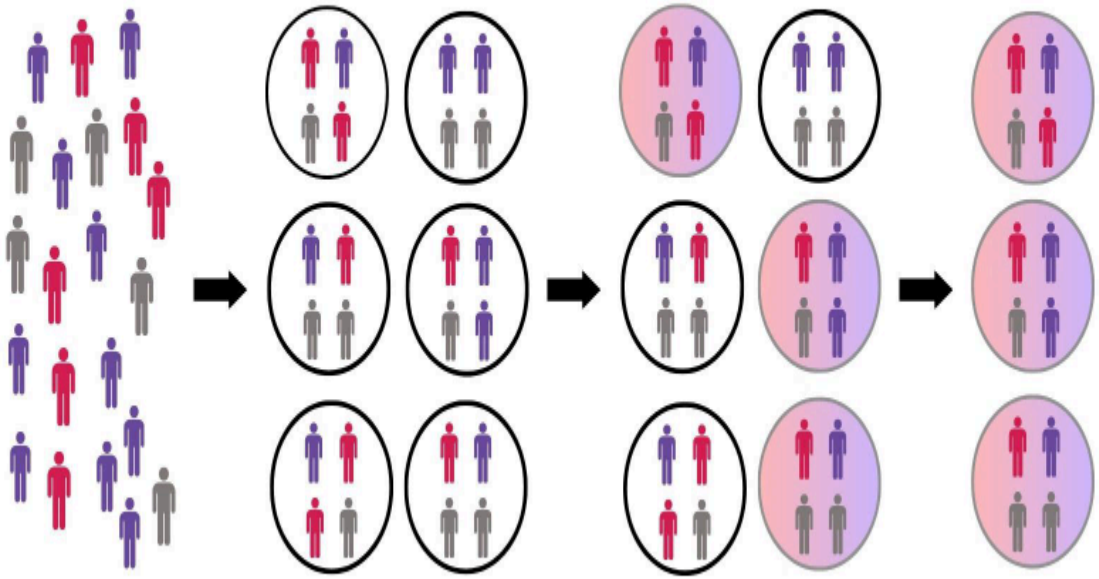
$$n_i = \frac{n \cdot N_i \sigma_i}{N_1\sigma_1 + N_2\sigma_2 + \ldots + N_k\sigma_k} \qquad \text{for } i = 1, 2, \ldots \text{ and } k.$$

We may illustrate the use of this by an example.

- **(iii) Cluster sampling:** If the total area of interest happens to be a big one, a convenient way in which a sample can be taken is to divide the area into a number of smaller non-overlapping areas and then to randomly select a number of these smaller areas (usually called clusters), with the ultimate sample consisting of all (or samples of) units in these small areas or clusters.
- Thus in cluster sampling the total population is divided into a number of relatively small subdivisions which are themselves clusters of still smaller units and then some of these clusters are randomly selected for inclusion in the overall sample.
- Suppose we want to estimate the proportion of machine parts in an inventory which are defective. Also assume that there are 20000 machine parts in the inventory at a given point of time, stored in 400 cases of 50 each.

TGM Research

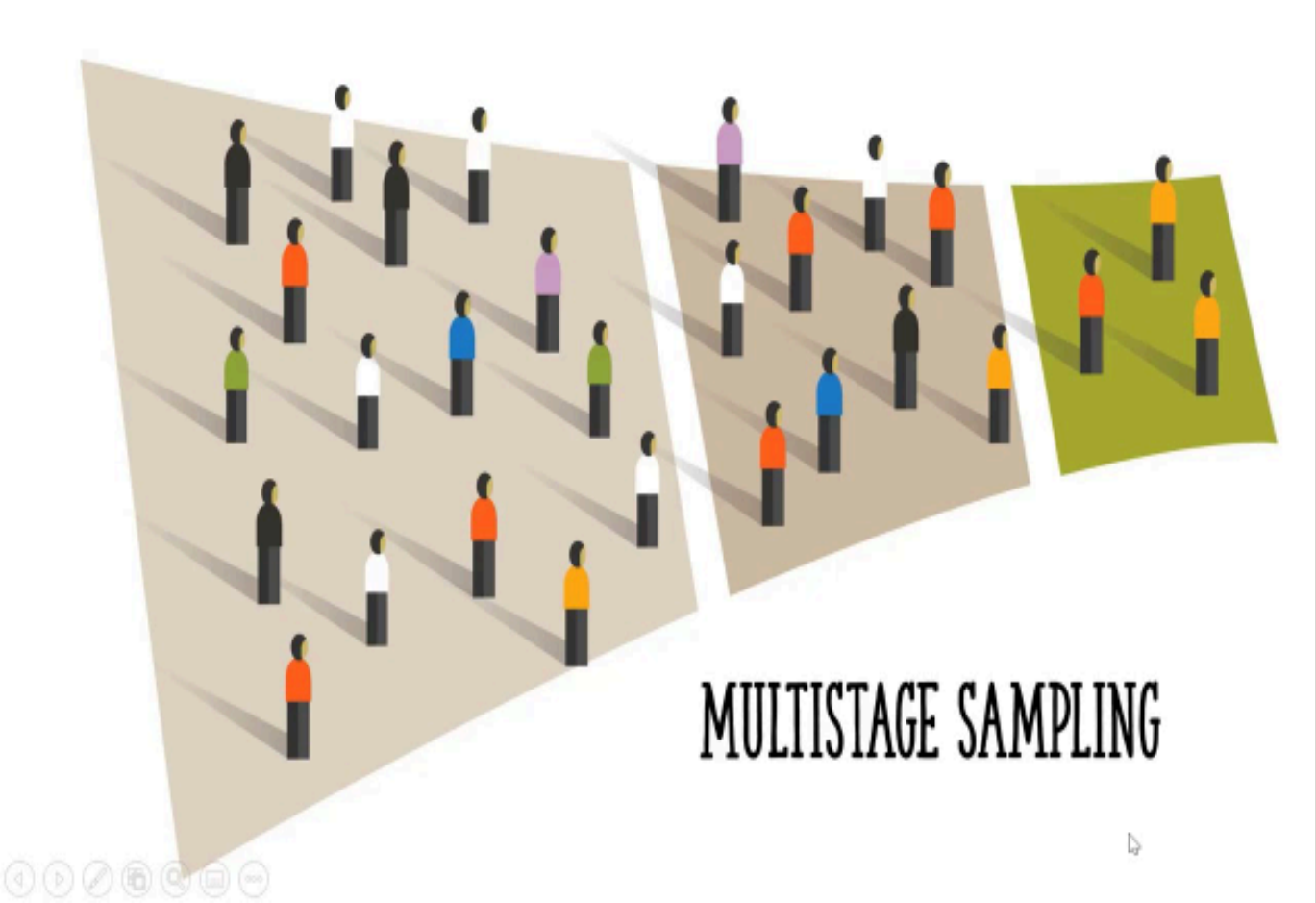
Cluster Sampling



- Now using a cluster sampling, we would consider the 400 cases as clusters and randomly select 'n' cases and examine all the machine parts in each randomly selected case.
- Cluster sampling, no doubt, reduces cost by concentrating surveys in selected clusters. But certainly, it is less precise than random sampling.
- There is also not as much information in 'n' observations within a cluster as there happens to be in 'n' randomly drawn observations.
- Cluster sampling is used only because of the economic advantage it possesses; estimates based on cluster samples are usually more reliable per unit cost.

(iv) Multi-stage sampling: Multi-stage sampling is a further development of the principle of cluster sampling.

- Suppose we want to investigate the working efficiency of nationalized banks in India and we want to take a sample of few banks for this purpose.
- The first stage is to select large primary sampling unit such as states in a country.
- Then we may select certain districts and interview all banks in the chosen districts.
- This would represent a two-stage sampling design with the ultimate sampling units being clusters of districts.
- If instead of taking a census of all banks within the selected districts, we select certain towns and interview all banks in the chosen towns.
- This would represent a three-stage sampling design.


MULTISTAGE SAMPLING

- If instead of taking a census of all banks within the selected towns, we randomly sample banks from each selected town, then it is a case of using a four-stage sampling plan.
- If we select randomly at all stages, we will have what is known as 'multi-stage random sampling design'.
- Ordinarily multi-stage sampling is applied in big inquires extending to a considerable large geographical area, say, the entire country. There are two advantages of this sampling design viz.,

(a) It is easier to administer than most single stage designs mainly because of the fact that sampling frame under multi-stage sampling is developed in partial units.

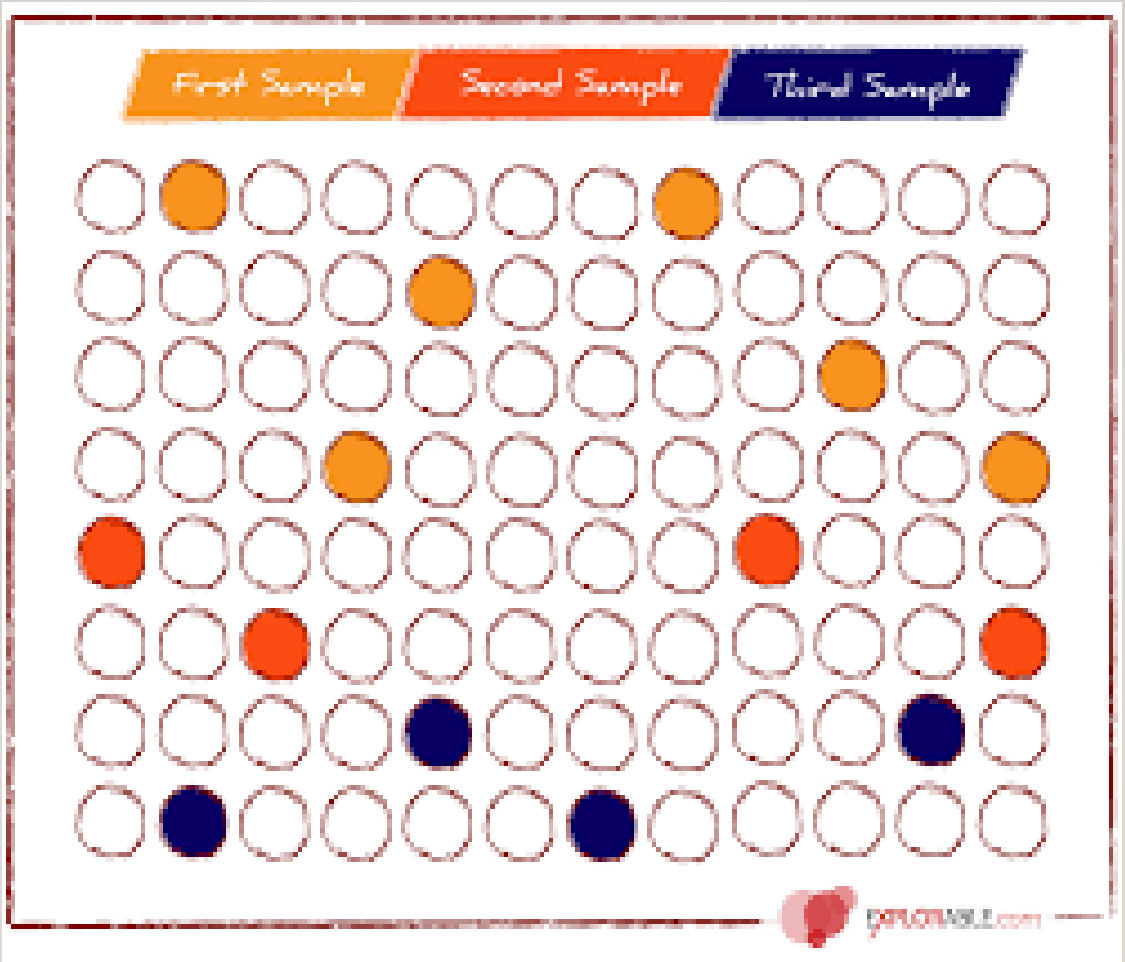
(b) A large number of units can be sampled for a given cost under multistage sampling because of sequential clustering, whereas this is not possible in most of the simple designs.

v) Sampling with probability proportional to size:

- In case the cluster sampling units do not have the same number or approximately the same number of elements, it is considered appropriate to use a random selection process where the probability of each cluster being included in the sample is proportional to the size of the cluster.
- For this purpose, we have to list the number of elements in each cluster irrespective of the method of ordering the cluster. Then we must sample systematically the appropriate number of elements from the cumulative totals.

- The actual numbers selected in this way do not refer to individual elements, but indicate which clusters and how many from the cluster are to be selected by simple random sampling or by systematic sampling.
- The results of this type of sampling are equivalent to those of a simple random sample and the method is less cumbersome and is also relatively less expensive.

- **(vii) Sequential sampling:**
- This sampling design is somewhat complex sample design.
- The ultimate size of the sample under this technique is not fixed in advance but is determined according to mathematical decision rules on the basis of information yielded as survey progresses.
- This is usually adopted in case of acceptance sampling plan in context of statistical quality control.


First Sample
Second Sample
Third Sample

- When a particular lot is to be accepted or rejected on the basis of a single sample, it is known as single sampling
- When the decision is to be taken on the basis of two samples, it is known as double sampling and in case the decision rests on the basis of more than two samples, but the number of samples is certain and decided in advance, the sampling is known as multiple sampling.
- But when the number of samples is more than two, but it is neither certain nor decided in advance, this type of system is often referred to as sequential sampling.
- Thus, in brief, we can say that in sequential sampling, one can go on taking samples one after another as long as one desires to do so.

MEASUREMENT AND SCALING

- **Introduction:** In our daily life we are said to measure when we use some yardstick to determine weight, height, or some other feature of a physical object.
- We also measure when we judge how well we like a song, a painting or the personalities of our friends.
- We, thus, measure physical objects as well as abstract concepts.
- Measurement is a relatively complex and demanding task, specially so when it concerns qualitative or abstract phenomena.
- Other examples of qualitative characteristics are taste, honesty, intelligence, and brand royalty.

QUANTITATIVE AND QUALITATIVE DATA

- Measurement is defined as a process of associating numbers or symbols to observations obtained in a research study.
- These observations could be quantitative or qualitative.
- For example, in case we are to find the male to female attendance ratio while conducting a study of persons who attend some show, then we may tabulate those who come to the show according to sex.
- In terms of set theory, this process is one of mapping the observed physical properties of those coming to the show (the domain) on to a sex classification (the range).

The rule of correspondence is:

- If the object in the domain appears to be male, assign to "0" and if female assign to "1".
- Similarly, we can record a person's marital status as 1, 2, 3 or 4, depending on whether the person is single, married, widowed or divorced.
- We can as well record "Yes or No" answers to a question as "0" and "1" (or as 1 and 2 or perhaps as 59 and 60).

- In this artificial or nominal way, categorical data (qualitative or descriptive) can be made into numerical data and if we thus code the various categories, we refer to the numbers we record as nominal data.
- Nominal data are numerical in name only, because they do not share any of the properties of the numbers we deal in ordinary arithmetic.
- For instance if we record marital status as 1, 2, 3, or 4 as stated above, we cannot write $4 > 2$ or $3 < 4$ and we cannot write $3 - 1 = 4 - 2$, $1 + 3 = 4$ or $4 / 2 = 2$.

- In those situations when we cannot do anything except set up inequalities, we refer to the data as ordinal data.
- For instance, if one mineral can scratch another, it receives a higher hardness number and on Mohs' scale the numbers from 1 to 10 are assigned respectively to talc, gypsum, calcite, fluorite, apatite, feldspar, quartz, topaz, sapphire and diamond.
- With these numbers we can write $5 > 2$ or $6 < 9$ as apatite is harder than gypsum and feldspar is softer than sapphire, but we cannot write for example $10 - 9 = 5 - 4$, because the difference in hardness between diamond and sapphire is actually much greater than that between apatite and fluorite.
- It would also be meaningless to say that topaz is twice as hard as fluorite simply because their respective hardness numbers on Mohs' scale are 8 and 4.
- The greater than symbol (i.e., >) in connection with ordinal data may be used to designate "happier than" "preferred to" and so on.